DOCUMENT RESUME

ED 205 542 TM 810 420

AUTHOR Popham, W. James
TITLE Development of Readability Controlled Basic Skills Tests.
PUB DATE Apr 81
NOTE 16p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basic Skills: *Basic Vocabulary: Elementary Secondary Education: Minimum Competency Testing: *Readability: Test Construction: Test Items: *Word Lists
IDENTIFIERS *Basic Skills Word List: South Carolina

ABSTRACT

Some minimum competency testing programs were created chiefly as a vehicle to improve the quality of basic skills instruction. In these programs, the tests are used to define instructional targets clearly and, thereby, enhance the quality of instruction. Few testing programs have included readability levels in their test specifications. In 1979 the staff of a test development agency was faced with the task of devising a set of basic skills tests for South Carolina which adhered to clearly defined readability levels. Since available readability formulas were inadequate for the creation of test items, a Basic Skills Word List was developed. These criteria were used for the development of a vocabulary pool: word frequency in reading texts, word frequency in general reading materials, and reader familiarity with particular words. Words were then allocated to grade levels proportionally, based on the number of words introduced in reading textbooks. The resulting Basic Skills Word List has both instructional and test-related applications. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

DEVELOPMENT OF READABILITY CONTROLLED BASIC SKILLS TESTS[1]

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. J. Popham

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

W. James Popham

University of California, Los Angeles

and

Instructional Objectives Exchange

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED205542

Educational achievement tests have historically functioned to provide an idea of a student's status with respect to a field of knowledge and skills, such as reading or mathematics. By surveying a broad range of skills and knowledge, achievement tests customarily yield estimates of an examinee's performance that are interpreted relatively according to the performances of other examinees in a norm group. Almost without exception, therefore, even if described as "diagnostic" assessment devices, most achievement tests have functioned as status-determination tools. Their chief mission has been to determine the status of students, typically at the close of some sort of instructional program.

Traditional norm-referenced achievement tests have also been employed to provide an estimate of the effectiveness of educational programs, such as a special instructional program in reading. Indeed, certain federal funding requirements almost oblige local educators to employ nationally normed assessment devices in their evaluative efforts.

TM 810 420

Basic Skills Tests as Goads

During the 1970s American educators focused considerable attention on the promotion of basic skills, that is, reading, writing, and mathematics. The "Back to Basics" movement was predicated on the belief

that insufficient instructional emphasis had been given to the basics, and that such a situation had to be rectified. The resurgence of interest in the basics was further stimulated by the adoption of minimum competency testing programs in almost 40 states. These testing programs, typically erected as antidotes to decreasing public confidence in the meaningfulness of high school diplomas, often required students to pass basic skills competency tests prior to graduation.

Although most of these minimum competency testing programs were installed to allay citizen incredulity regarding the quality of schooling, some minimum competency testing programs were created chiefly as vehicles to improve the quality of basic skills instruction. Such an improvement orientation is clearly the case in the legislation establishing South Carolina's Basic Skills Assessment Program and the Texas Assessment of Basic Skills. The High School Proficiency Program of the Detroit Public Schools is another instance in which a basic skills testing program was established primarily to improve instruction.

In these improvement-oriented testing programs we encounter achievement tests which still function as status-determination mechanisms (that is, in which students can display proficiency in the basic skills). But, more importantly, in these programs the tests are used primarily as vehicles to define instructional targets clearly and, thereby, enhance the quality of instruction by permitting teachers to design and implement instructional sequences which are more on-target. Putting it another way, because the tests attempt to stake out with more clarity the nature of the basic skills to be promoted, teachers are thereby encouraged to provide students with more on-target practice opportunities, that is, more "time-on-task."

These new types of basic skills tests, therefore, have taken on a function not served by their predecessors, namely, the improvement of instruction. More than serving as mere status-determination mechanisms, this new genre of basic skills tests functions as a spur to instructional improvement.

In order for a basis skills test to stimulate instructional improvement, it is imperative for that test to explicate with considerable precision the nature of the skills being measured by the test. In fact, it is the lucidity with which the test's skills are defined that permits the test to illuminate the skills being sought, hence enhance teachers' comprehension of the skills they must promote.

Test Specifications as Illuminators

Recent work in the field of criterion-referenced measurement has emphasized the key communicative role played by the test's specifications, that is, the rules employed to generate the actual items used on the test. Sometimes referred to as "domain specifications," since those specifications operationalize the domain of behaviors being assessed by the test, the test's specifications provide teachers with the skill definitions needed to organize their instruction. The more lucid that such specifications are, the more likely that educators will understand the skill being sought. The better that educators understand the skill being sought, the more likely that they will design appropriate instructional sequences.

During the past decade there have been various approaches employed in the construction of criterion-referenced test specifications (Hambleton, 1978). These approaches range in the degree to which they circumscribe

eligible test items, some providing far more restrictions in the the content, format, and wording of test items than others. Other than at a very general level, no criterion-referenced test specifications have attacked the readability level of the test items. Yet, the readability of the test's items plays a vital role in clarifying the nature of the skills to be tested.

At best, some test developers have attempted to employ traditional readability formulae to constrain the reading level of test items. But these formulae were developed for use with extensive written passages, not with the brief sentences and phrases often used in objective test items. Beyond that, there are some substantial shortcomings with readability formulae if one's intent is to clarify rigorously the nature of the skill being tested by explicating the nature of the items measuring the skill.

Procedures for determining the readability levels of written passages have been available for a number of years. Most of these procedures are based on quantifiable factors such as the numbers of words in sentences and the numbers of syllables in words (e.g., Flesch, 1948; Fry, 1968). These sorts of readability formulae usually do not take into consideration a reader's actual familiarity with the words being rated. To illustrate, imagine a very short sentence consisting exclusively of one-syllable, yet obscure words. Since the sentence is brief and the words are short, its readability level as determined by most readability formulae would be low. Yet it may present a difficult reading task even for very skilled readers. Conversely, one can conceive of a fairly lengthy sentence composed of polysyllabic but very familiar words. Such a sentence, although its readability level is high, would represent a

fairly easy reading task for most readers. It is apparent that to get an accurate idea of the readability of a given selection we must attend not only to such structural features as sentence length and number of syllables, but also to the words themselves. Most readability formulas offer little guidance in this area.[2]

The Need for a Basic Skills Word List

In 1979 the staff of a test development agency was faced with the task of devising a set of basic skills tests in reading, writing, and mathematics for the state of South Carolina. These tests were supposed to adhere to clearly defined readability levels. Members of the test development staff discovered that available readability formulas were clearly inadequate for the creation of test items which were at a reading level unequivocally suitable for students at a specific grade level. As indicated earlier, most readability formulae can be applied only to fairly extensive reading selections. Even when grade-by-grade constraints on sentence length and syntactic complexity were set, it was impossible to tie grade level readability unless test developers also relied on a word list. But word lists based on different strategies yield different sets of words. Which word list was to be used?

There are three different sources which have been utilized as the basis for word lists. These are (1) the frequency of words appearing in published reading textbooks series, (2) the frequency of words appearing in generally read materials, e.g., newspapers, magazines, and books, and (3) readers' reported or tested familiarity with particular words.

Typically, a word list is prepared on a grade-by-grade basis using one of these three strategies. In each approach the assumption is that

words more frequently encountered by individuals (or more well known) will be more appropriate at lower grade levels. Yet, although there will obviously be overlap in word lists based on these three approaches, there will also be substantial differences among the word lists generated by relying on each of these three strategies.

The test development staff was not obliged to choose only one approach from among the three strategies, that is, (1) word frequency in reading texts, (2) word frequency in general reading materials, and (3) reader familiarity with particular words. Fortunately, three recently compiled word lists reflecting each of these three strategies were available. By combining these three separate words lists, it was possible to assemble a fundamental vocabulary which simultaneously reflected all three criteria. Since its chief use was to be the generation of a set of basic skills tests, the new vocabulary list was designated as the Basic Skills Word List.

Development of the Basic Skills Word List

In creating the Basic Skills Word List a series of separate steps were followed in order to create a vocabulary pool which would systematically reflect three criteria. The initial source of words was the EDL Core Vocabularies (Taylor, S., et al., 1979). This widely used volume contains word lists for each grade from pre-primer through 13. These lists (particularly at the lower grades) are based on the frequency with which words appear in nine widely used basal reading textbook series. Typically, if a word appeared in three or more of the textbook series at a given grade level, it was included in the EDL list at that grade level. Thus, this set of words is based on the frequency of usage in reading textbooks.

All the words on the EDL list were checked for their familiarity to children by using Dale and O'Rourke's study, The Living Word Vocabulary--The Words We Know (Dale and O'Rourke, 1976). The authors of this volume determined students' knowledge of commonly encountered words by administering multiple-choice test items to students. Students were given a word and asked to choose the correct definition for it. The Dale-O'Rourke vocabulary provides a "familiarity percentage" for each word listed. This index reflects the percentage of children who answered that word's multiple-choice test item correctly. In order to assign a word to a particular grade level, the authors of the study aimed for each tested word to have a familiarity percentage for a given grade level that was within the range of 67%-84%. Therefore, if a word was tested at the sixth grade and only 66% of the students at that grade were familiar with it, then the word was re-tested at the eighth grade. Conversely, if a word tested at the eighth grade received a familiarity score higher than 84%, it was re-tested at the sixth grade. The familiarity percentage supplied with each word in the Dale-O'Rourke vocabulary reflects the percentage of correct student responses at the grade level to which the word was ultimately assigned.

The authors of the The Living Word Vocabulary did not begin testing words until the fourth grade and, after that, tested words only at alternate grades. Thus, familiarity percentages appear only for grades 4, 6, 8, etc. Therefore, EDL words in grades 1-4 were checked for their familiarity to students according to the fourth grade Dale-O'Rourke familiarity percentages. Fifth grade EDL words were checked against both fourth and sixth grade familiarity percentages. Words in all subsequent grades' lists were checked for familiarity ratings at either

the grade level at which they appeared in EDL or at a lower grade.

Words that were not familiar to at least 65% of students at a given grade on the basis of the Dale-O'Rourke study were moved to a higher grade level in the Basic Skills Word List. The exact familiarity percentages necessary for an EDL word to be retained at the same grade level on the Basic Skills Word List varied slightly from grade to grade. These percentages were adjusted in order to meet the requirements of a pre-determined word load for each grade. This word load factor will be described subsquently.

The rationale for employing a stringent familiarity criterion was straightforward, namely, that even if a word is found in several reading series at a given grade level, it may still be unfamiliar to many children and therefore should not be assigned to that grade level. The effect of this student familiarity screen was to move some words from each of the graded EDL word lists to higher grade levels.

Grade-Level Word Load

For instructional purposes, it is desirable to allocate words to grade levels on a proportional basis. It would make little sense to assign 200 words at one grade level and 2,000 words to another. One of the best guides to the determination of an appropriate word load per grade level is the average number of words introduced per grade level by publishers of reading textbook series. These commercially published textbooks, many of them revised more than once, provide an experience-based estimate of how many new words can be reasonably introduced at each grade level. In the process of researching the background for their core vocabularies, developers of the EDL word lists calculated the

average number of words introduced at grades once through six for nine different textbook series. These textbook-derived word loads were the following:

Grade	1	2	3	4	5	6
Word Load	341	440	708	787	1,063	1,077

The word loads for the Basic Skills Word List at grades one through four were designed to coincide as closely as possible with these textbook-derived word loads. The word loads for the Basic Skills Word List at grades one through four are the following:

Grade	1	2	3	4
Word Load	341	439	708	785

In grades 5-12, students' familiarity with words, as reflected by the Dale-O'Rourke study, became a major determinant of grade level word load for the Basic Skills Word List. Students at these grade levels displayed insufficient familiarity with many potentially eligible words, thus reducing the word loads--particularly in grades 9-12. The word loads for the new Basic Skills Word List in grades 5-8 are approximately 900 per grade. In grades 9-12 the approximate word load is 400. The word loads for grades 5-12 follow:

Grade	5	6	7	8	9	10	11	12
Word Load	971	846	884	874	325	407	393	345

The final step in the selection of words for the Basic Skills Word List was based on a massive study by Carroll, Davies, and Richman (Carroll, J., et al., 1971). This study analyzed 5,000,000 running words of text. These 5,000,000 words were taken from approximately

10,000 samples of 500 words excerpted from textbooks in 17 different curriculum areas in grades 3-9, plus magazines, books, newspapers, and poetry. The result of the study is a list of 87,000 words, accompanied by the frequency with which each of those words appears in print. Unfortunately, this enormous set of words is listed alphabetically, rather than in the order of each word's frequency of usage. Hence, one cannot readily determine the most frequently encountered words.

Recently, however, Sakiey and Fry have drawn on the 87,000 words to provide 3,000 Instant Words (Sakiey, E. and Fry, B., 1979), a list of the 3,000 most frequently occuring words ranked according to their usage in print. Words from 3,000 Instant Words, in order of decreasing frequency of appearance, were added at each grade level of the Basic Skills Word List if they were not already listed. This insured that words that appear very frequently in general reading materials were not overlooked when they did not have a high enough familiarity percentage.

Word List Usage Rules

In general, only the root form of a word has been listed in the Basic Skills Word List. The derived form of a word is listed separately only if it is more commonly used than the root form or if its meaning is significantly different from the root form. Proper names were excluded, as were most proper nouns and sound or movement words such as "meow" or "zoom." Provincialisms and colloquialisms were also omitted. No attempt was made to include technical vocabularies for the social, physical, or biological sciences. Common contractions and abbreviations have been included.

When a word appears on the list it is to be taken as its most common usage. At grade levels higher than its listing, a word may be

employed in other than its most common usages and as another part of speech.

For teachers and other individuals who want to use the Basic Skills Word List to prepare instructional materials, a list of usage rules has been compiled. These rules indicate the types of changes that can be made to the words listed at each grade level. All the rules for a grade level also apply to all subsequent grade levels. The use of these rules can be illustrated by considering the first-grade rules. If a noun is listed as a grade one word and its plural is formed by adding "s," then the noun's plural form can also be used at that grade level. For example, "boy" is a grade one word. Therefore both "boy" and "boys" are eligible for use at grade one (and all other grade levels). These rules accompany the published version of the new word list (IOX, 1980).

Instructional Applications

Sets of fundamental vocabulary terms such as the Basic Skills Word List provide grade-by-grade terms which pupils need to master. The more defensibly that those word lists were devised, of course, the better. Having access to such sets of grade-designated words permits educators to diagnose students' word knowledge strengths and weaknesses in a systematic manner. Vocabulary diagnostic exercises can be based on particular grade-level word lists so that teachers can identify students who need additional vocabulary-building instruction.

Such word lists also provide a set of words to be fostered in classroom vocabulary-building activities. Teachers can focus their efforts on promoting student mastery of a basic vocabulary judged to be suitable at the student's own grade level. Remedial instruction can

deal with sets of words designated as appropriate for earlier grade levels.

Teachers can also use these sorts of word lists as a tool to gauge the readability of instructional materials, either those distributed by commerical publishers or those developed locally. If such materials appear to reflect a vocabulary level not consonant with the grade level at which the teacher is teaching, then the grade level of questionable words can be quickly ascertained through the use of such word lists.

Use of the Word List in Test-Related Settings

Teachers can employ such word lists as definitive vocabulary guides which can prove valuable for purposes of instructional design or the development of tests. In situations where the basic skills tests that are being employed have been constructed according to the requirements of a particular word list, teachers will have a clear idea of vocabulary constraints placed on test items. By consulting the readability limits placed on test items for a particular grade level, teachers can get a precise fix on all the eligible words which can be used in test items for their students. During instruction, therefore, teachers can stress those particular words so that students will become conversant with the full range of eligible words.

In settings where no tests based on a particular word list have been prescribed, educators may wish to create their own tests which rely on a given word list, thus delimiting the eligible sets of words that students should master.

Clearly, the optimum dividends to be secured from use of any basic skills word list arise when testing is coupled with teaching. Tests

which are carefully constructed to reflect specified readability levels provide teachers with defensible instructional targets since all words that are "fair game" will have been identified. As a consequence, teachers who promote their pupils' familiarity with the stipulated sets of words will be giving those pupils an optimum opportunity for success on the tests. A student's mastery of a particular intellectual skill will not be obscured by a test item's use of terms unfamiliar to that student. An equitable testing system will have been created. More importantly, perhaps, instruction will have been installed which is attuned to the test instruments that are employed to assess that instruction's effectiveness.

References

Carroll, John, Peter Davies, and Barry Richman, Word Frequency Book, American Heritage Publishing Co., Inc., New York, 1971.

Dale, Edgar and O'Rourke, Joseph, The Living Word Vocabulary--The Words We Krow, Dome Press, Elgin, Illinois, 1976.

Flesch, Rudolf, "A New Readability Yardstick," Journal of Applied Psychology, Vol. 32, No. 3, June 1948, 221-223.

Fry, Edward, "A Readability Formula That Saves Time," Journal of Reading, Vol. 11, April 1968, 513-516, 575-578.

Hambleton, R. K. et al. Criterion-Referenced Testing and Measurement: A Review of Technical Issues and Development, Review of Educational Research, Vol. 48, No. 1, Winter 1978, 1-48.

IOX, IOX Basic Skills Word List, Grade 1-12, IOX, Los Angeles, 1980, Box 24095, Los Angeles, CA 90024

Sakiey, Elizabeth and Edward Fry, 3,000 Instant Words, Jamestown Publishers, Providence, R.I., 1979.

Taylor, Standard, et al., EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies, McGraw-Hill, New York, 1979.

X11WJP/K

Footnotes

[1]A symposium presentation at the annual meeting of the American Educational Research Association, Los Angeles, April 13-17, 1981.

[2]One widely known readability procedure, the Dale-Chall formula, does incorporate a word familiarity feature. The list of familiar words which accompanies the formula was compiled by having fouth-graders check the words they knew from a list of 10,000 words. The resulting word list is now well over 30 years out of date. For a description of the readability scheme devised by Dale and Chall, a procedure quite revolutionary for its time, see Dale, Edgar, and Chall, Jeanne S., "A Formula for Predicting Readability," Educational Research Bulletin, vol. 27, January 21, 1948, pp. 11-28.